A Comparison of two Sampling Designs for use in the B.C. Forest Inventory: Stratified Probability Proportional to Size with Replacement (PPSWR) and Multiple Pass Ordered Systematic (MPOS) Sampling.

Peter Ott
Biometrician
B.C. Forest Service
Research Branch
Victoria, B.C.
V8W 9C2

July 1999

106-04940/

Introduction

On February 2, 1999, a meeting was held at Resources Inventory Branch to discuss the report by Sit and Ott (1999), which reviewed the current B.C. Forest Inventory audit sampling scheme and estimation methods. A central concern that arose in the discussion was the lack of an unbiased estimate of the estimated variance (of the mean) under the current ordered systematic (OS) design[1]. It was agreed that this feature alone makes the OS sampling design unsatisfactory. Some participants in the discussion proposed as an alternative a stratified probability proportional to size with replacement (PPSWR) design[2], whereas others advocated a multiple pass OS design, claiming that the benefits of a systematic design are too important to be discarded. In light of these points, the Director of Resources Inventory Branch requested that Research Branch conduct an investigation to compare the performance of the stratified PPSWR and multiple OS designs. This report compares the two designs using simulated samples from the Lillooet TSA.

The MPOS design is conducted by taking *m* OS samples from the population (i.e. the entire list of polygons), each with independent random starts. This type of design is often referred to as 'replicated sampling' (Deming 1956, SAS Institute Inc. 1999) or 'the random groups technique' (Särndal et. al. 1992). The stratified PPSWR design involves selecting polygons with replacement from predetermined strata with probability proportional to their area.

Analysis objectives

The objective of this study is to compare the performance of estimators of the mean and variance (of the estimated mean) for the MPOS and PPSWR designs using (a) volume per ha, and to a lesser extent (b) age and height.

Analysis procedure

The population used for our study was the Lillooet TSA. It contains 45,715 forest inventory polygons, 37,727 of which are treed (Table 1). The dataset was presorted by the

[1] The origin of the OS design can be traced back to Madow (1949).
[2] This design was first developed by Hansen and Hurwitz (1943).

following criteria: non-vegetated/vegetated, non-treed/treed, leading species, age, and site index. Please refer to J.S. Thrower & Associates (1999) for further details. Two sets of analyses were conducted. One used only treed polygons as the population, and the other used the entire landbase.

Table 1. Population breakdown of Lillooet TSA.

Land Classification	Number of Polygons	Area (ha)	Proportion of Total Area
Non-vegetated and vegetated non-treed	7,988	478,205.96	42.5
Treed	37.727	646,203.48	57.5
Total	45,715	1,124,409.44	100

Each population was sampled using (a) a stratified PPSWR design and (b) a suite of multiple pass OS (MPOS) designs with fixed total sample size (i.e. 3 passes of 40, 6 passes of 20, and 12 passes of 10). For both designs, 1000 independent samples were drawn and estimator performance was investigated.

We only examined simple single stage estimators of mean volume per ha, age and height (and their associated variances). The simulations assume that the population values (e.g. volume per ha) in each polygon are precisely known, which is obviously not the case in reality. In other words, we did not attempt to include photo-interpreter error. However, the simplified approach will help clarify the magnitude of sampling error.

Within-polygon sampling was not conducted since the polygons are assumed to be homogeneous within. Hence, two stage estimators for both designs are equivalent to the one stage estimators. We chose not to impose any artificial within-polygon-variability since it would (a) affect estimators form both designs in an equal manner, (b) complicate the results, and (c) add considerable processing time.

The pre-assignment of polygons to strata was designed to roughly match the sort criteria used for the MPOS design. For example, considering the treed population only, we divided the population into groups defined by leading species and three broad age classes (age < 100, $100 \leq$ age < 200, and age ≥ 200). For each of the initial strata, we allocated

samples proportional to areal representation. From there, we collapsed together alike but sparsely sampled strata until at least one sample was available to each stratum. In the end, rather than dropping a stratum with only one sample, we 'took' one sample from each of the two most heavily sampled strata and 'gave' these to it (giving it three samples). The resulting pre-stratified arrangement is shown in Table 2.

Table 2. The 9 different strata used for the PPSWR sample of the Lillooet TSA (treed polygons only). Polygons ≥ 200 years are denoted as 'very old', and those <100 years are denoted as 'young'. The stratum defined as 'Miscellaneous' consists of hemlock, red cedar, and hardwood stands. The last column indicates the number of samples taken within each stratum for the PPSWR design.

Strata	Number of Polygons	Area (ha)	Proportion of Total Area	Number of Samples
Douglas-fir very old	4,324	83,946.04	13	16
Douglas-fir old	6,690	126,539.04	20	22
Douglas-fir young	1,874	28,216.64	4	5
Pine very old	2,309	45,009.92	7	8
Pine old	6,776	122,496.92	19	22
Pine young	3,426	63,783.24	10	12
True fir	6,663	97,373.64	15	18
Spruce	5,234	73,461.16	11	14
Miscellaneous	431	5,376.88	1	3
Total	37,727	646,203.48	100	120

A similar process was used to designate strata for the entire landbase. Since the non-vegetated and vegetated non-treed polygons consistently have zero volume, we set aside a single stratum with merely 3 samples for this portion of the population (43%). We then allocated the remaining 120–3=117 samples to the rest of the landbase as before. The results are shown in Table 3.

Table 3. The 10 different strata used for the PPSWR sample of the Lillooet TSA (all polygons). Explanation of terminology can be found in the title of Table 2.

Strata	Number of Polygons	Area (ha)	Proportion of Total Area	Number of Samples
Douglas-fir very old	4,324	83,946.04	7	15
Douglas-fir old	6,690	126,539.04	11	22
Douglas-fir young	1,874	28,216.64	3	5
Pine very old	2,309	45,009.92	4	8
Pine old	6,776	122,496.92	11	21
Pine young	3,426	63,783.24	6	12
True fir	6,663	97,373.64	9	18
Spruce	5,234	73,461.16	7	13
Miscellaneous	431	5,376.88	0.5	3
Non-vegetated and vegetated non-treed	7,988	478,205.96	43	3
Total	45,715	1,124,409.44	100	120

Results and discussion

In all analyses, none of the MPOS passes encountered polygons with areas larger than the selection interval k. For example, in the treed population the smallest observed value of k (i.e. when 40 samples per pass of were taken) was 16,155.09 and the largest polygon was only 2,465.44 ha. The smallest interval used to sample the entire landbase was 28,110.24 and the largest polygon was 11,438.88 ha.

Volume per ha

The results using volume per ha suggest that (as expected) both methods yield unbiased estimates of the mean (Table 4a). Estimated variances (of the estimated mean) should also be unbiased in theory, but our results (Table 4b) deviated slightly due probably to the moderate number of simulations relative to the small sampling fraction. Regardless, it appears that the estimated variances are approximately unbiased, and that the true

variances for the MPOS are slightly lower than the PPSWR, except for the case of 12 passes of 10 samples.

Table 4a. Simulation results for estimated mean volume per ha. Population consists of treed polygons only (N=37,727).

Method / Estimator	Number of Passes	Number of Samples per Pass	True Mean	Estimated mean	Relative Bias (%)
PPSWR	1 (9 strata)	120	134.64	135.13	0.36
MPOS	3	40	134.64	134.65	0.01
MPOS	6	20	134.64	134.42	-0.16
MPOS	12	10	134.64	135.18	0.40

Table 4b. Simulation results for estimated variance of mean volume per ha. Population consists of treed polygons only (N=37,727).

Method / Estimator	Number of Passes	Number of Samples per Pass	True Variance	Estimated Variance	Relative Bias (%)
PPSWR	1 (9 strata)	120	86.48	88.32	2.13
MPOS	3	40	72.67	67.47	-7.71
MPOS	6	20	74.80	81.33	8.03
MPOS	12	10	92.38	94.69	2.43

Age and Height

The results using age (Tables 5a and 5b) and height (Tables 6a and 6b) are similar to those for volume per ha except that the true variance observed for the PPSWR sample is substantially larger than that from the MPOS samples. This is due to the fact that both age and height are nearly inversely proportional to area for this population, and hence the PPSWR sample is actually significantly less efficient than a SRS or an OS sample. For example, a SRSWR sample of this data using age yielded a true variance of 19.72.

The inefficiency of the PPSWR design for age and height is not crucial since these variables are generally analysed using ratio estimation techniques. Furthermore, the ratio

of the sampled age (or height) with its photo interpreted representation is a much more critical attribute in the inventory since it has more direct importance to timber supply projections, and timber harvest scheduling.

Table 5a. Simulation results for estimated mean age. Population consists of treed polygons only (*N*=37,727).

Method / Estimator	Number of Passes	Number of Samples per Pass	True Mean	Estimated mean	Relative Bias (%)
PPSWR	1 (9 strata)	120	160.40	157.84	-1.60
MPOS	3	40	160.40	160.40	0.003
MPOS	6	20	160.40	160.30	-0.06
MPOS	12	10	160.40	160.56	0.10

Table 5b. Simulation results for estimated variance of mean age. Population consists of treed polygons only (*N*=37,727).

Method / Estimator	Number of Passes	Number of Samples per Pass	True Variance	Estimated Variance	Relative Bias (%)
PPSWR	1 (9 strata)	120	509.05	507.46	-0.31
MPOS	3	40	8.87	8.98	1.32
MPOS	6	20	8.54	8.62	0.88
MPOS	12	10	17.16	15.68	-8.61

Table 6a. Simulation results for estimated mean height. Population consists of f treed polygons only (*N*=37,727).

Method / Estimator	Number of Passes	Number of Samples per Pass	True Mean	Estimated mean	Relative Bias (%)
PPSWR	1 (9 strata)	120	18.78	18.63	-0.80
MPOS	3	40	18.78	18.78	0.02
MPOS	6	20	18.78	18.77	-0.05
MPOS	12	10	18.78	18.79	0.05

Table 6b. Simulation results for estimated variance of mean height. Population consists of treed polygons only (N=37,727).

Method / Estimator	Number of Passes	Number of Samples per Pass	True Variance	Estimated Variance	Relative Bias (%)
PPSWR	1 (9 strata)	120	6.99	6.99	-0.01
MPOS	3	40	0.21	0.21	-0.19
MPOS	6	20	0.28	0.27	-3.86
MPOS	12	10	0.28	0.27	-1.95

Table 7. Comparison of the distribution of samples for the 3-pass MPOS (40 samples per pass) with the 9-strata PPSWR sample allocation for the Lillooet TSA (treed polygons only). Data in the first three columns are based on the combined number of samples falling into each category for each simulation. The last column indicates the (predetermined) number of samples taken for the PPSWR design.

Strata	Median Number of Samples (out of 1000)	Minimum Number of Samples	Maximum Number of Samples	Number of Samples used for PPSWR
Douglas-fir very old	15	15	18	16
Douglas-fir old	24	21	24	22
Douglas-fir young	5	3	6	5
Pine very old	8	6	9	8
Pine old	23	21	24	22
Pine young	12	10	12	12
True fir	18	18	20	18
Spruce	14	12	15	14
Miscellaneous	1	0	3	3
Total	120			120

Distribution of samples under the MPOS

How consistently does the MPOS sample encompass the breadth of desired characteristics for the population? We addressed this question by examining the combined number of samples that fell into predetermined strata. We chose to use the same strata already defined for the PPSWR design since these were analogous to how the population was sorted. Table 7 shows that the range of samples that fall into each stratum is quite small, implying that a single sample (multiple pass) should sensibly encompass the population.

Number of Strata

The impact of changing the number of strata was explored by (1) first reducing the number of strata from nine in Table 2 to five by disregarding age and using only leading species to delineate the strata, and (2) then collapsing the resulting five strata from (1) into only three strata: Douglas-fir, Pine and 'everything else'. As expected, reducing the number of strata did not affect the bias in the estimators but it did lead to an increase in the true variance. Specifically, the variance increased from 86.48 to 95.69, and then to 105.61 as the number of strata decreased.

Entire Landbase

As with the treed population, the estimates of the mean and variance are unbiased (Tables 8a and 8b). However, by reducing the number of samples in the non-vegetated and vegetated non-treed areas, we substantially improved the efficiency of the PPSWR design over the MPOS design (Table 8b). The flexibility of the sample placement is one benefit of the stratified PPSWR design.

Table 8a. Simulation results for estimated mean volume per ha. Population consists of the entire landbase (N=45,715).

Method / Estimator	Number of Passes	Number of Samples per Pass	True Mean	Estimated mean	Relative Bias (%)
PPSWR	1 (10 strata)	120	77.39	76.95	-0.58
MPOS	3	40	77.39	77.56	0.22
MPOS	6	20	77.39	77.31	-0.10
MPOS	12	10	77.39	77.28	0.13

Table 8b. Simulation results for estimated variance of mean volume per ha. Population consists of the entire landbase (N=45,715).

Method / Estimator	Number of Passes	Number of Samples per Pass	True Variance	Estimated Variance	Relative Bias (%)
PPSWR	1 (10 strata)	120	29.38	30.18	2.72
MPOS	3	40	51.88	53.22	2.58
MPOS	6	20	62.04	63.64	2.59
MPOS	12	10	75.18	75.28	-0.15

We believe that the trends for age and height noted in the treed population will extend to the entire landbase in a similar fashion to volume per ha.

Implementation Considerations

The discussion so far has chiefly centred on the accuracy (i.e. bias and precision) of the two sampling designs. However, there are other factors that should be considered in deciding which one to implement. Listed below are some potential challenges along with speculative proposed solutions.

1. How does one select extra polygons, after the initial sampling selection has been made?

This is trivial for the PPSWR design if the original list is maintained. Simply make another random selection (with probability proportional to polygon area) from the list. For execution of the MPOS design, it is important that independent passes, each with new random starts are utilised. Appendix C discusses another less appealing approach. It seems prudent to keep the original list in the event that more passes are required at a later date, rather than initially obtaining extra passes (samples). This avoids the temptation to snoop through the set of extra passes for samples having particular properties.

2. What does one do when the chosen polygon does not meet the requirements for the population (e.g. a selected stand turns out to be logged, but only stands older than 60 years are required to be included in the target population)?

This is again trivial for the PPSWR design. One would ignore the invalid selection, remove it from the population, and then redraw another polygon. The obvious option for the MPOS design would be to simply ignore the selection and remove it from the population. The possibility of each pass not having the same number of sampled polygons is not a major concern. A related matter is the situation where the selected polygon is physically inaccessible. Although somewhat biased, one possible solution for both designs may be to substitute the selection with a polygon that is believed to be similar (and pretend that it was the one originally selected).

3. Is it possible to post-stratify the sample(s) and still obtain unbiased estimates? Also, is it straightforward to pre-stratify the MPOS design?

Although it would be quite unusual to take a pre-stratified PPSWR sample and then decide to post-stratify the sample with different criteria, it would be possible to obtain unbiased parameter estimates if one keeps the original list and sample (especially the polygon areas). The MPOS design can be pre-stratified or post-stratified. Estimation would involve first calculating the totals for the attributes of interest for each stratum, separately for each pass. Next, the mean and variance of each stratum total would be calculated using the information from the multiple passes. Finally, the mean and associated variance for the entire unit would be calculated by summing over the estimates from each stratum and then dividing by the total area (for the mean) and the squared total area (for the variance of the mean).

4. Is it possible to combine samples from two subunits, even though one may overlap with the other? For example, a complete sample is taken from a TSA which overlaps with a TFL. Can the TFL use for part of its sample those polygons which were selected in the TSA sample?

This can be easily accomplished in the PPSWR design by treating the TSA as a unique stratum, and then sampling the remainder of the TFL as per usual. For the each pass of the MPOS design, every hectare has an equal chance of being selected which is the inverse of the selection interval (n / *total area*). Therefore a single pass of the TSA can be matched with a single pass from the remainder of the TFL as long as each polygon is weighted by a factor of n_k / *total area*$_k$ before combining them (k is used here to index multiple disjoint areas). After the samples from the two areas have been successfully concatenated for each pass (using the weights), then the estimates for total area can be calculated in the usual way.

5. Is it possible to use sampling information based on an earlier inventory to assess the characteristics of an updated inventory? In other words, even though the target population from which the sample was obtained has changed, is there a valid way to manipulate the sample so that it reflects the new population?

Since the population has changed, it will be impossible to use previous sampling information to describe the exact ground attributes of the updated inventory. One could proceed by artificially 'growing' the sample obtained at the earlier date (using a man-made model) but this will potentially introduce bias and imprecision. For the sake of discussion, let us assume that this bias and imprecision is negligible. The polygons will need to be reconstituted — after the new line work is overlaid, cluster plots falling outside the new polygon boundary should be omitted and plots that fell outside the original polygon boundary (but that are now inside the new boundary) should be recovered and measured. These new clusters must then be recompiled and factitiously grown to describe the alleged ground characteristics of the new polygons. At this point, the new polygon areas would be needed to apply the PPSWR estimators. For the MPOS design, the above exercise needs to be conducted for each pass, and then the new means from each pass would be used to estimate the overall mean and its variance.

6. In a two step VRI implementation process, how should the final ground sample be controlled? Usually, the initial sample size estimate is transitory and the final sample

size is computed using the coefficient of variation from intermediate ground sampling information. For example, if the final sample size turns out to be smaller than the initial guess, can a multiple pass sample be truncated to suit the new sample size specifications?

Increasing or decreasing the initial sample size is obvious for the PPSWR design. It becomes more difficult for the MPOS design since sampling effort consists of two components: (a) the number of passes, and (b) the number of polygons per pass. A determination of sample size should clearly specify both (a) and (b). Since it is probably easier to increase or decrease the number of passes (rather than the number of polygons per pass), it seems sensible to use this as the transitory parameter. This avoids the prospect of having to modify the number of sampling units in the final pass.

Conclusions

Shiue (1960) showed that systematic sampling with multiple random starts is equivalent to cluster sampling (where each cluster consists of a set of elements distributed evenly over the population). This result does not extend to the MPOS design because the population cannot be factored into $a \times b$ (with a being the maximum number of passes) uniquely sampled elements (i.e. the samples are not necessarily disjoint). This is due to the probability proportional to size component inherent in the MPOS scheme. Regardless, the implication of this is minor and we may still use the estimated means from each pass to obtain an unbiased estimate of the variance of the overall mean.

It appears that the MPOS design adequately yields unbiased estimates and a low true variance. Further work may be required to determine the most optimum appropriation of sampling effort; is it better to increase the number of passes or the number of samples per pass? The main drawback to the MPOS design is the additional effort involved, especially in intricate situations such as trying to combine samples from different areas or time periods. Not only is more work involved than the PPSWR design, but there is an increased chance for error.

Based on the results of the application of the PPSWR and MPOS sampling designs to the Lillooet population data, the estimators for the designs have desirable characteristics for different reasons. In terms of bias in variance estimation, the MPOS is marginally superior to the PPSWR in that it can provide a smaller true variance. From the

perspective of practical application, PPSWR would be less demanding on data management effort, and would lend itself more easily to adding more VRI samples in the future. With the MPOS design, the potential for making major data processing errors over the lifetime of the inventory (20 to 30 years) is considerable. Tracking multiple passes, and weights for three or more subunits within a district over a long period of time could become unnecessarily arduous. Even with the PPSWR design, there would be a need to track subunit weights, but there would only be one sample dataset for each unit or subunit.

Acknowledgements

Thanks to Sam Otukol, Gary Johansen and Darrell Errico. This report would not have been possible without their continual input and technical support.

References

Cochran, W. G. 1977. Sampling Techniques, Third Edition. John Wiley & Sons Inc., NY.

Deming, W. E. 1956. On simplifications of sampling design through replication with equal probabilities and without stages. Journal of the American Statistical Association. 51:24-33.

Hansen, M. H. and W. N. Hurwitz. 1943. On the theory of sampling from finite populations. Annals of Mathematical Statistics. 14:333-362.

J. S. Thrower & Associates Ltd. 1999. Lillooet Forest District - VRI Ground Sampling Plan (Revised). BC Min. Forests, Victoria, BC.

Madow, W. G. 1949. On the theory of systematic sampling, II. Annals of Mathematical Statistics. 20:333-354.

Särndal, C.-E., Swensson, B. and J. Wretman. 1992. Model Assisted Survey Sampling. Springer-Verlag. NY.

SAS Institute Inc. 1999. SAS/STAT Software: Experimental Procedures in the Initial Release of Version 7 of the SAS System (Draft). Cary, NC. Chapter 6. http://www.sas.com/rnd/app/da/new/dasurvey.html.

Sit, V. and P. K. Ott. 1999. A Statistical Review of the B.C. Forest Inventory Audit Sampling Scheme and Estimation Methods. BC Min. Forests, Victoria, BC.

Shiue, C.-J. 1960. Systematic sampling with multiple random starts. Forest Science. 62:1:42-50.

Thompson, S. K. 1992. Sampling. John Wiley & Sons Inc., NY.

Notation used in the Appendices

MPOS - ordered systematic sampling

PPSWR - stratified probability proportional to size with replacement

z_i - area of polygon i

Z - total area of all polygons in population

Z_h - total area of polygons in stratum h

y_i - audit volume per hectare for polygon i

Y_i - total audit volume for polygon $i = y_i z_i$

n - number of polygons sampled

n_h - number of polygons sampled from stratum h

N - total number of polygons in population

N_h - total number of polygons in stratum h

k - sampling interval = Z/n

m - total number of passes (samples), indexed by $l = 1, 2, \ldots, m$.

L - total number of strata

$\hat{y}_{NAME}$ - estimate of mean volume per ha

$v(\bullet)$ - estimated variance of the estimate

$\hat{y}_s$ - value of estimator for simulation s

S - total number of simulations

Appendix A. Estimation formulae for the MPOS and stratified PPSWR estimators

The following estimators can be also found in textbooks such as Cochran (1977) or Thompson (1992). Note that all of these are single stage estimators (i.e. within-polygon variability is not considered).

Most of what follows is presented in terms of the variable: volume per ha. Parameter estimates for the variables age and height are almost identical (with the exclusion of the 'per area' component in the formulae) and hence, are not discussed.

Stratified PPSWR

The stratified PPSWR estimate of mean volume per ha is given by

$$\hat{y}_{PPSWR} = \frac{1}{Z}\sum_{h=1}^{L}\hat{Y}_h \quad \text{where} \quad \hat{Y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h}\frac{Y_i}{p_i} \quad \text{and} \quad p_i = \frac{z_i}{Z_h} .$$

The estimated variance of this estimate is

$$v\left(\hat{y}_{PPSWR}\right) = \frac{1}{Z^2}\sum_{h=1}^{L} v\left(\hat{Y}_h\right) \quad \text{where} \quad v\left(\hat{Y}_h\right) = \sum_{i=1}^{n_h}\frac{\left(\frac{Y_i}{p_i} - \hat{Y}_h\right)^2}{n_h\left(n_h - 1\right)} .$$

MPOS

The MPOS estimate of mean volume per ha is given by

$$\hat{y}_{MPOS} = \frac{1}{m}\sum_{l=1}^{m}\hat{y}_l \quad \text{where} \quad \hat{y}_l = \frac{1}{n}\sum_{i=1}^{n} y_i .$$

The estimated variance of this estimate is

$$v\left(\hat{y}_{MPOS}\right) = \sum_{l=1}^{m}\frac{\left(\hat{y}_l - \hat{y}_{MPOS}\right)^2}{m(m-1)} .$$

Appendix B. Simulation method

The true population mean volume per ha is given by $y = \frac{\sum_{i=1}^{N} Y_i}{Z}$.

and the estimated mean is $\bar{\hat{y}} = \frac{\sum_{s=1}^{S} \hat{y}_s}{S}$.

Therefore, relative bias for the mean is calculated using $RB = 100 \cdot \left(\frac{\bar{\hat{y}} - y}{y} \right)$.

The true variance is calculated using the mean squared error: $MSE = \frac{\sum_{s=1}^{S} \left(\hat{y}_s - \bar{\hat{y}} \right)^2}{S-1}$.

The estimated variance is $\overline{v(\hat{y})} = \frac{\sum_{s=1}^{S} v_s(\hat{y})}{S}$,

and so the relative bias in the variance is given by $RB = 100 \cdot \left(\frac{\overline{v(\hat{y})} - MSE}{MSE} \right)$.

For the simulation study of age and height, analogous calculations to these were used (i.e. we used '*y*' to denote mean age or height rather than mean volume per ha) with one exception: the true mean age or height was calculated using an area weighted mean. For example, the calculation of true mean age using the above notation, would look something like:

$$\textit{true mean age} = \text{'}y\text{'} \text{ (in above notation)} = \frac{\sum_{i=1}^{N} z_i \cdot age_i}{Z}.$$

Appendix C. Sampling from a Large Initial Sample

A scenario might arise where a large OS sample of size n_1 has been obtained, and we wish to obtain *m* equal sized OS samples, each of size n_2 ($n_2 <<< n_1$) using the larger sample as a template. By conditioning the second set of samples on an initial large sample, this approach heavily restricts the random starting point used for each of the individual passes. The statistical properties of this system were not investigated but Särndal et. al (1992) have shown that the estimator of variance is biased (p. 426). Therefore, it should only be performed as a way to salvage older samples.

To begin with, the second sample should be chosen so that n_1/n_2 is an integer[3], and the original sample should neither be randomised nor stratified. If we denote the interval length for the original systematic sample as k_1 (units are in ha), then by definition, $k_1 = \frac{total\ area\ of\ population}{n_1}$.

The interval size for each of the second set of m samples can be expressed as a function of k_1. Specifically, $k_2 = \frac{n_1}{n_2} k_1$.

The fraction n_1/n_2 indicates the number of polygons (in the original sample) that must be jumped to obtain one of the second samples, and still properly encompass the population. It also represents the maximum number of passes allowed. For example, if n_1/n_2=50, then we would turn to the original sample, and choose polygons from it commencing at some arbitrary polygon. If we began with the first polygon from the sample, the new sample would consist of polygons 1, 51, 101,..., 1+(n_2-1)·50. However, there is no strong justification for choosing the first polygon from the first sample, and there are obviously multiple approaches for deciding where to start (and continue) the selection process. Here are a few possibilities:

1. We start with the first selected polygon from the first sample (as in the example above). From there, obtain a sample of polygons 1, $1+n_1/n_2$,..., $1+(n_2-1)\cdot n_1/n_2$. Next, start with the second selected polygon from the first sample. From there, obtain a sample having polygons 2, $2+n_1/n_2$,..., $2+(n_2-1)\cdot n_1/n_2$. We continue this process until we reach the mth polygon from the first sample; m must be $\leq n_1/n_2$. We use it to select a sample of polygons m, $m+n_1/n_2$,..., $m+(n_2-1)\cdot n_1/n_2$. This gives us m passes of size n_2.

2. Another approach might be to choose the starting polygon by making a random selection among the first n_1/n_2 polygons in the first sample. Once this polygon has been chosen, continue using its adjacent polygons as the rest of the starting points and sample similar to (1) above until m sets of samples are obtained.

3. A third approach would be to select each of the m starting points completely randomly (without replacement) from the first n_1/n_2 polygons chosen in the first sample.

---

[3] If this fraction is not an integer, then round it *down* to the nearest integer. However, this will affect the degree to which the sample encompasses the entire distribution of polygons.